Contents lists available at ScienceDirect

Safety and Health at Work

journal homepage: www.e-shaw.org

Original Article

Development of a Tailored Analysis System for Korean Working Conditions Survey

CrossMark

Hwa Jeong Seo

Medical Informatics and health Technology (MIT), Department of Healthcare Management, College of Social Science, Gachon University, Seongnam, Republic of Korea

ABSTRACT

Background: Korean Working Conditions Surveys (KWCS), referencing European Working Conditions Surveys, have been conducted three times in order to survey working condition and develop work-related policies. However, we found three limitations for managing the collected KWCS data: (1) there was no computerized system for managing data; (2) statistical KWCS data were provided by limited one-way communication; and (3) the concept of a one-time provision of information was pursued. We suggest a web-based public service system that enables ordinary people to make greater use of the KWCS data, which can be managed constantly in the future.
Methods: After considering data characteristics, we designed a database, which was able to have the result of all pairwise combinations with two extracted data to construct an analysis system. Using the data of the social network for each user, the tailored analysis system was developed. This system was developed with three methods: clustering and classification for building a social network, and an infographic method for improving readability through a friendly user interface.
Results: We developed a database including one input entity consisting of the sociodemographic characteristics and one output entity consisting of working condition characteristics, such as working pattern and work satisfaction. A web-based public service system to provide tailored contents was completed.
Conclusion: This study aimed to present a customized analysis system to use the KWCS data efficiently, provide a large amount of data in a form that can give users a better understanding, and lay the ground for helping researchers and policy makers understand the characteristics.

Copyright © 2016, Occupational Safety and Health Research Institute. Published by Elsevier. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The Korean Working Conditions Survey (KWCS) is planned to identify workers' exposure to working factors by investigating the working conditions of Korean workers nationwide. Looking into the history of KWCS, the KWCS for workers was conducted as a nationwide sample survey to contribute to industrial health and safety policy establishment for the improvement of working conditions in 2006 [1]. In 2010 [2] and 2011 [3], the KWCS was also carried out to collect additional data necessary for changes and policy decision making on working conditions. The following limitations in storing and managing the data collected from the KWCS have been recognized as obstacles to efficient KWCS data utilization.

First, there is no computerized system to manage the data from the KWCS. The only method that uses the KWCS data currently involves downloading the processed data via the homepage of Statistics Korea. However, not many people know that Statistics Korea provides KWCS data; although users know where they can get the KWCS data, ordinary users who are not experts in analysis or interpretation may have difficulty in understanding the meanings of these data and in gaining knowledge because these are mass basic data that have gone through cleansing. Second, most interactives applying statistical technique to open data are limited to one-way communication, which provides precalculated results for the users to select from; it is difficult to induce users' participation and constant attention through one-way communication [4] based on a service provider's design without considering users' need. In

Medical Informatics and health Technology (MIT), Department of Healthcare Management, College of Social Science, Gachon University, 1342 Seongnamdaero, Sujeong-gu, Seongnam, Gyeonggi 461701, Republic of Korea.
E-mail addresses: hjseo@gachon.ac.kr; sharryseo@gmail.com.

contrast, a two-way communication system can perform an analysis of all possible combinations of variables, allowing diverse exploration according to users' interests. For example, some users may wonder about the employment contract type for people under the same conditions (occupation, region, etc.) as theirs, while others may wonder about the health conditions of individuals whose age is similar to theirs, including their smoking or drinking rate. Current services cannot reflect such requirements. Third, it is important to go beyond one-time provision of information. The KWCS data aim to be open to general users, as well as to researchers and policy-makers; satisfy their right to know; and provide useful information. However, it is difficult to expect general users' constant participation and continuous use of the current KWCS data due to limited information and the absence of optimum service to arouse interest.

This study then aims to suggest a web-based public service system that enables ordinary people to make greater use of the KWCS data, which can be managed constantly in the future. To achieve the main goal of our study, two objectives were created: the first was to design a database and establish an analysis system so that general users can be provided with the contents of analysis based on the KWCS data, and the second was to present an analysis algorithm that can reflect users' needs through a two-way communication system [5].

2. Materials and methods

2.1. Data acquisition

2.1.1. Data source
The first phase data of the KWCS in 2006 analyzed in this study included 10,043 cases [1], the second phase data in 2010 included 10,019 cases, and the third phase data in 2011 included 50,032 cases [3].

2.1.2. Data processing
Characteristics of the KWCS data were classified into sociodemographic and working conditions. The sociodemographic characteristics consist of five items—gender, age, industrial type, occupation type, and region—and the working condition characteristics consist of six items—labor structure, working pattern, work satisfaction, work environment, and the health and nutrition indices [6]. On the basis of this data classification, a database schema was drawn.

The database was established as shown in Fig. 1. First, the basic structure was classified into sociodemographic items and working condition items from the raw data. Second, we defined and identified the input and output entities. Data from sociodemographic items were difficult to analyze because the data were either continuous with an undefined range or discrete with a lot of permissible values. Thus, they were clustered on the basis of similarity among data. For example, the typical continuous data "age" was divided into three clusters—< 30 years, 30–39 years, and 50 years or older—while the permissible values of discrete data were divided into 10 or fewer categories. They constituted an input entity. By contrast, for the KWCS data of working condition items, there was a well-defined classification system for the values of discrete data. Thus, they were directly reflected in the design. For example, the labor structure entity included properties such as employment contract type and continuous working period. They constituted an output entity. Finally, databases were designed and built as having data with all pairwise combinations between the sociodemographic characteristics (input entity) and the working environment characteristics (output entity).

We applied a clustering technique for the major items of the questionnaire, including age, gender, service period, income,
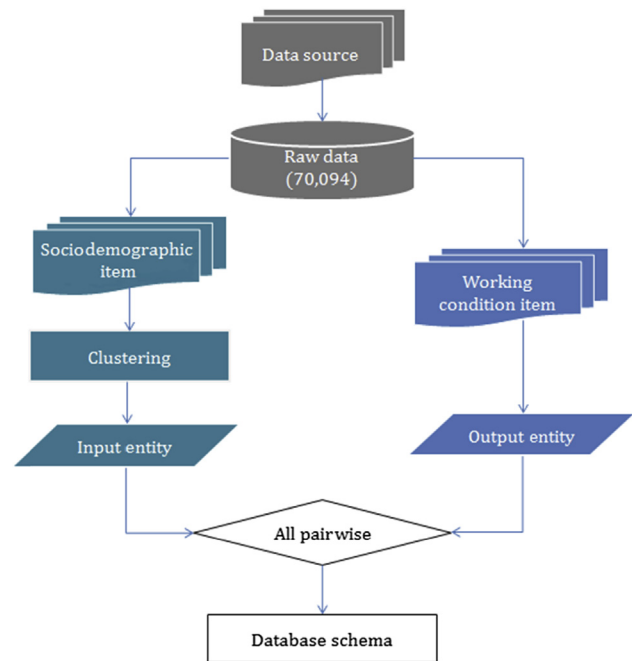

Fig. 1. Flowchart for developing the database schema.

region, industrial classification, employment contract type, and labor hours. The classification technique was also applied to guide new KWCS data into the constructed database. In other words, while clustering serves to be conducted for grouping in the given data, classification is a technique used to judge to which group new data or information will belong when the new data or information comes in from the grouping classification system [7,8]. The purpose of classification, based on the user's (new) inputted information, must be to obtain information on user's similar group and reference (control group) (Fig. 2).

2.1.3. Data representation
Information visualization (or infographic) [9,10] is a methodology used to express data that contain certain information or knowledge, and to show the data efficiently and conveniently. Using graphical factors, infographics aims to express and efficiently convey information to users so that significance can be extracted from the information. Our study suggested to users how to apply a dress code, which was friendly and more readable through "avatars" [11,12] for reflecting themselves. In other words, we used this infographic method to transform users' input (gender, age, occupation type, etc.) into information. The dress code of users' avatar was changed according to the input to set "my own appearance." For example, if a user wants to represent a man of < 30 years in the industrial category of financing and insurance as an office worker, the mascot is a green avatar (which means a male) in the dress code of a protective helmet with one logo (which means having an age of < 30 years), and in a white shirt (which means an office worker) [13].

2.2. Analysis algorithm

Johari's window is a conflict analysis method based on communication. It examines conflict causes on the basis of communication in a relationship between the user and others [14,15]. Following this method, the two-way communication proposed in this study functions in a way where conflicts (problems) are solved by receiving feedback from others after exposing oneself
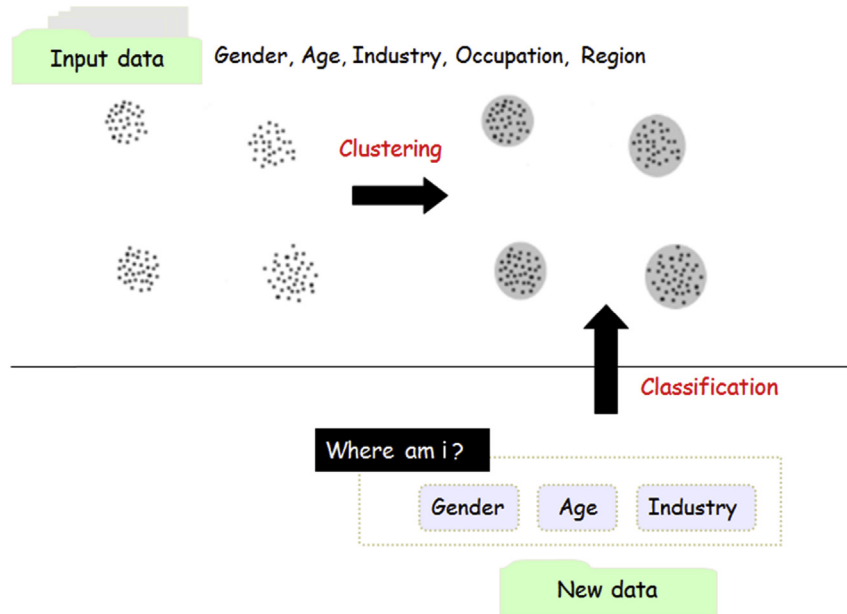
Fig. 2. Clustering and classification according to demographic data.

to the others (Fig. 3). The concepts of social network and tailored analysis algorithm for analyzing user information are also adopted with the data collected from a similar group.

2.2.1. Similar and comparison groups within a social network

A social network means a cluster of networks formed on the basis of the basic data that users provide. This has been drawn from the meaning of a similar group, which is a group of data with great similarity among them; that is, a cluster (clustering) of groups identical or similar to oneself is a similar group, or "one's own appearance." In contrast, a control group is a cluster of groups other than similar groups; that is, a group independent of similar ones or a cluster of groups other than similar ones is a control group.

2.2.2. Tailored analysis algorithm

The analysis algorithm in this study aims to make a comparative analysis of working conditions between one's social network, which is a group identical to one's own, and a social network, which is a control group. This method enables the support of interactive contents to analyze the properties that a user wants, and the formation of social support among members of the same social network or people in the same situation [16,17].

3. Results

3.1. Input and output entities

3.1.1. Input entity and its properties for sociodemographic items

This study has established a social network that enables one to provide minimum information and receive feedback on the basis of

the information. To form a social network that could give feedback, clustering was carried out on the basis of the individuals' gender, age, industry, and so on.

We clustered five unspecified data items including gender, age, industry, occupation and region from sociodemographic characteristics, to complete the input entity and its properties. Gender was fragmented into male and female; age was fragmented into younger than 30 years, 30−49 years, and 50 years or older; and industry was fragmented into mining, manufacturing, electricity, gas and waterworks, construction, transportation, warehousing and communications, agriculture, forestry and fishery, finance and insurance, and others. Occupation was fragmented into office workers and production workers, and region was fragmented into Seoul metropolitan area, Chungcheong region, Gyeongsang region, Honam region, and others. In this study, which cluster the concerned data belong to was resolved based on the user-inputted data (Table 1).

3.1.2. Output entity and its properties for working condition items

As a result of analyzing the surveyed results in three occasions with reference to European Working Conditions Surveys [18,19] data, the KWCS items were classified into six categories: (1) labor and force structure; (2) work pattern; (3) work satisfaction; (4) working conditions; (5) health impact index; and (6) accident, disease, and other experiences. The details of each item are presented below (Table 2). Labor force structure is classified into employment contract type and period of work. Work pattern is clustered into the time period covering the time of coming to work and leaving work, while satisfaction with work is classified into satisfaction with working conditions and job continuity. Work
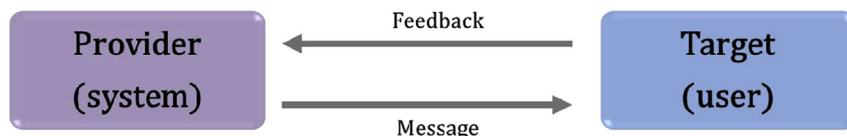

Fig. 3. Two-way communication functions.

Table 1
Division type of input entity and its properties

Entity	Property		Value
Input	Gender	1	Male
		2	Female
	Age (y)	1	Under 30
		2	30–49
		3	50 and above
	Industry	1	Mining
		2	Manufacturing
		3	Electricity, gas and waterworks
		4	Construction, transportation, warehousing and communications
		5	Agriculture, forestry and fishery
		6	Finance and insurance
		7	Others
	Occupation	1	Office workers
		2	Production workers
	Region	1	Seoul metropolitan area
		2	Chungcheong region
		3	Gyeongsang region
		4	Honam region
		5	Others

condition is clustered into stages of physical risk factors, while health impact index is clustered into health problem occurrence, smoking amount, and drinking frequency. Accident, disease, and other experiences are classified into absence arising from work-

Table 2
Division type of output entity and property

Entity	Property		Value
Labor force structure	Employment contract type	1	Regular workers
		2	Nonregular workers
		3	Others
	Period of work	1	Less than 5 y
		2	5–10 y
		3	10 y or more
Work pattern	Rush hours	1	Less than 1 h
		2	1–2 h
		3	2 h or more
		4	None of the above
		5	Don't know/nonresponse
Satisfaction with work	Satisfaction with work conditions	1	Satisfied
		2	Unsatisfied
		3	Don't know/nonresponse
	Sustainable jobs	1	Yes
		2	No
		3	Don't want to answer
		4	Don't know/nonresponse
Working condition	Risk factors for physical work	1	More than average
		2	Average
		3	Less than average
Health impact index	Health problem occurrence	1	Yes
		2	No
	Smoking amount	1	Never smoke
		2	Less than five packs of cigarettes
		3	More than five packs of cigarettes
	Drinking frequency	1	More than four times a week
		2	Two or three times a week
		3	Two or four times a month
		4	Less than once a month
		5	Never drink
Accident, disease, and other experiences	Absenteeism due to occupational accidents	1	Yes
		2	No
		3	Don't know/nonresponse
	Absenteeism due to work-related diseases	1	Yes
		2	No
		3	Don't know/nonresponse

related accident, absence due to work-related disease, and unpleasant experience for the past 1 year (i.e., outcast/bullying and sexual harassment).

3.1.3. Entity relationship diagram for KWCS database
We generated an entity relationship model [20] for representing the refined KWCS data using entity relationship modeling, which is the primary method for database design (Fig. 4). Our entity relationship model is designed for referencing two entities: (1) input entity including one table (see Table 1) and (2) output entity including six tables (see Table 2).

3.2. Tailored analysis system

3.2.1. Results of analysis according to input entity
As for the respondents' employment contract type in 2011, there were 29,148 men and 20,884 women. Of the men, 58.0% were regular workers, 12.5% were nonregular workers, and 29.5% belonged to the others group. Of the women, 50.8% were regular workers, 17.4% were nonregular workers, and 31.8% belonged to the others group. The employment contract type differed statistically significantly by gender ($p < 0.05$).

Information on "employment contract type by gender" in this study was obtained on the basis of "row %" (differences in employment contract type) and not column % (differences in gender of respondents). As for employment contract type by gender, 61.4% were men and 38.6% women, and men were more likely to be employed on a regular basis than women (Table 3). This information is provided to users infographically.

3.2.2. User's image and comparison group
With a few scenarios, we will explain how a "man of less than 30 years in the industrial category of financing and insurance and the occupational category of office work in the region of Gyeongsang Province" forms his social network and obtains information via the tailored analysis system.

A user might want to obtain information as a result of the queries "in what employment contract type people of the same gender as me (similar group) work and what percentage do they form?" and "in what employment contract type people of a certain gender compared with me (control group) work and what percentage do they form?" Since "I" is male in a social network, the employment contract type and percentage for males (similar group) and the social group-gender (females) as a control group were employed. As for the employment contract type by gender of the respondents in 2011, more males (61.4%), including me, were found to work as regular workers compared with females (38.6%) in the control group (Fig. 5).

The website address of the tailored analysis system that we developed is as follows: http://www.kosha.or.kr/jsp/kwcs/index.do?fw=index&menuId=7598.

3.2.3. Dress code of an avatar
In the social network of the abovementioned user, the avatar of a male office worker of < 30 years is a green mascot (male) in the dress code of a protective helmet with one logo (< 30 years) and a white shirt (office worker). Fig. 6A shows a dress code of an avatar for a male office worker of < 30 years and Fig. 6B shows an avatar in a dress code for a blue-collar woman aged 30 years to < 50 years.

4. Discussion

The KWCS was created to identify exposure to working factors such as occupation, industry, and employment type through the survey of working conditions of Korean employees nationwide, by
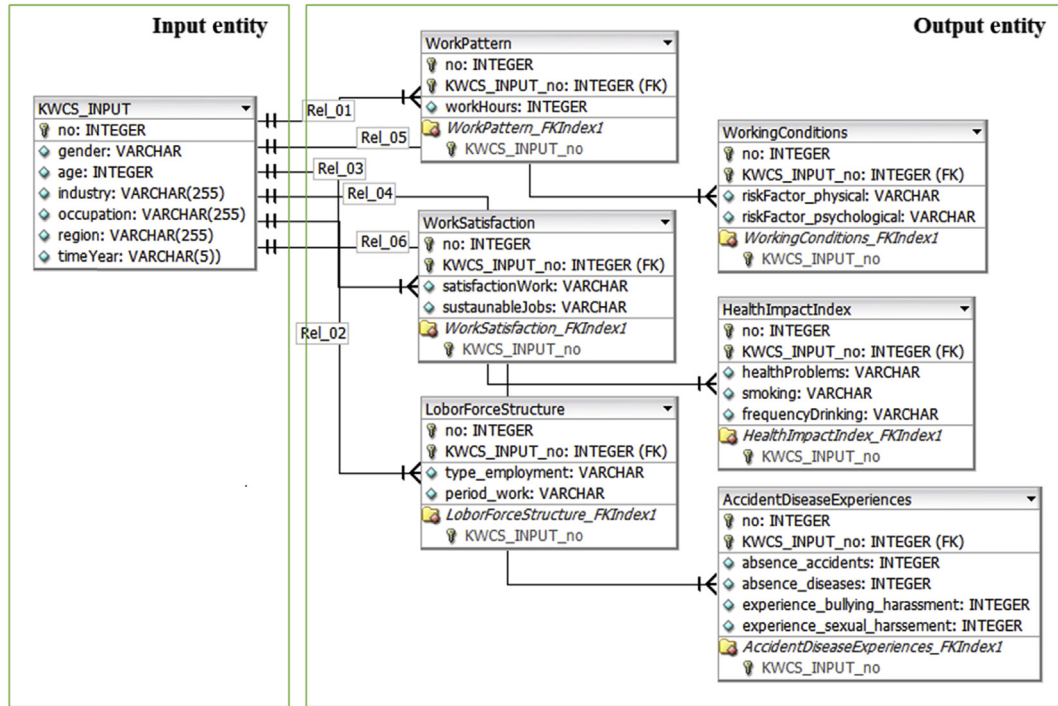
Fig. 4. ER diagram for the KWCS database. ER, entity relationship; KWCS, Korean Working Conditions Survey.

referring to the European Working Conditions Surveys. Based on the planning concerned, KWCS for employees was conducted with a nationwide sample survey in 2006, in an effort to contribute to the establishment of industrial health and safety policies for the improvement of working conditions. The second and third phases of the KWCS were carried out to collect the data necessary for the identification of changes and policy decision making on working conditions in 2010 and 2011, respectively. The surveys conducted thus far aimed to identify the impacts of sociopsychological factors, including safety, health, employment type, and job stress, as well as mechanical, physical, and chemical risk factors, on working conditions. The surveys also aimed to improve working conditions to reduce health risk factors in a working environment, not just limiting the aim to simple data collection and situation identification.

This study aimed to present a customized analysis system to use the KWCS data efficiently, provide a large amount of data that are regularly produced in a form that could give users a better understanding, and lay a ground for helping researchers and policy makers understand characteristics, such as working types, according to various working conditions. Our tailored analysis system for the KWCS data draws the following conclusions on the basis of the ultimate goal to realize informatization of user convenience-oriented working conditions:

A two-way communication analysis system has been established to present the results of analysis of the KWCS data in 2006, 2010, and 2011 as inputted by users, including data on gender, age, industry, occupation, and region. This system has applied a control group presentation algorithm based on a social network to attract the attention of people and majors in general. Clustering is carried out according to sociodemographic variables such as gender, age, industry, occupation, and region. The results for a control group against the target group (my appearance) are presented when new data are inputted. An analysis system similar to this one has been developed for the European Working Conditions Survey. While clicking a country on a map visualizes a comparison with the European mean or gives expression of data in a proper bar graph and table, it has a limitation of being based on one-way communication that only shows data already defined by the researchers. Our study is expected to be useful in reflecting a user's needs by designing and developing an analysis system based on two-way communication, to avoid the above-mentioned issue.

Table 3
Employment contact type by gender

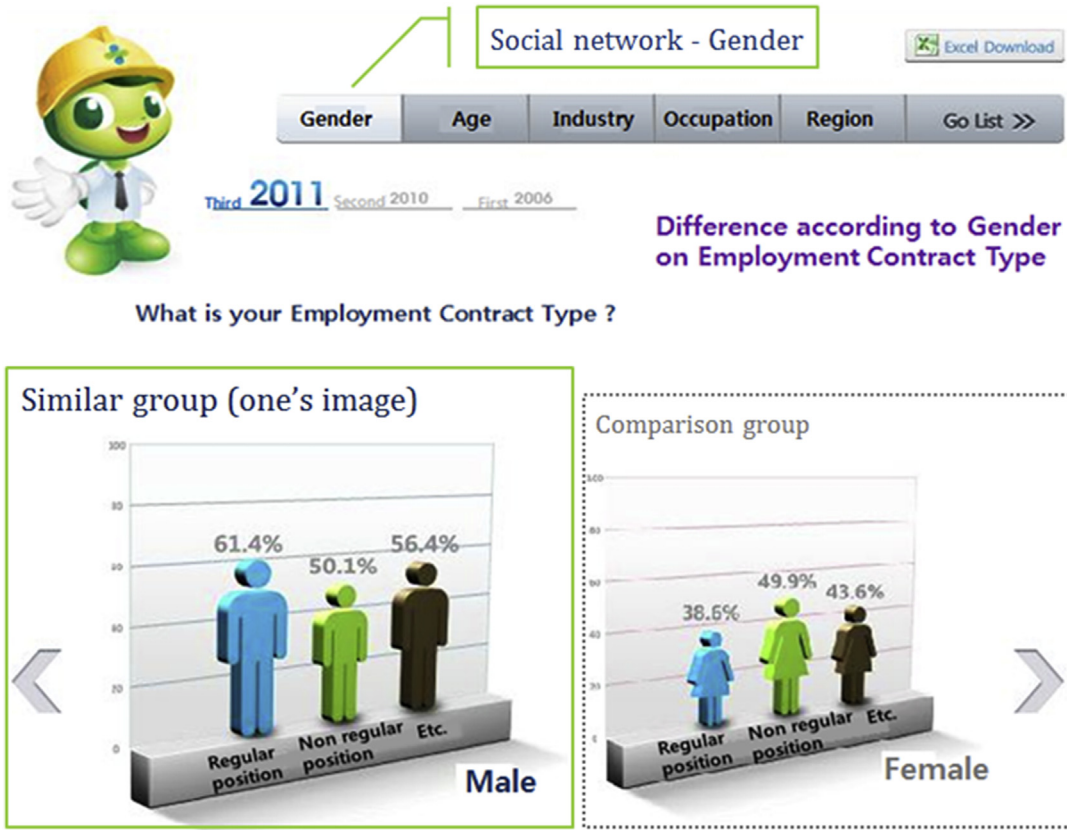|		Employment contact type			Total	x^2	p
		Regular	Nonregular	Others			
Gender							
Male	n	16,893	3,650	8,605	29,148	336.262	0.000
	Column (%)	58.0	12.5	29.5	100		
	Row (%)	61.4	50.1	56.4	58.3		
Female	n	10,600	3,638	6,646	20,884		
	Column (%)	50.8	17.4	31.8	100		
	Row (%)	38.6	49.9	43.6	41.7		
Total	n	27,493	7,288	15,251	50,032		
	(%)	100	100	100	100		

Fig. 5. Similar group (one's image) and comparison group in a social network (gender).

Second, "intuitive visualization" was used instead of tabulation for simple representation in expressing analysis data. It was from the lesson of the case of National Institute of Occupational Safety and Health (NIOSH). NIOSH under the control of Centers for Disease Control and Prevention (CDC) has established a lead exposure database. This database system is used to present lead exposure database available in public health communities in simple graphs and tables on the basis of the data collected individually in a region
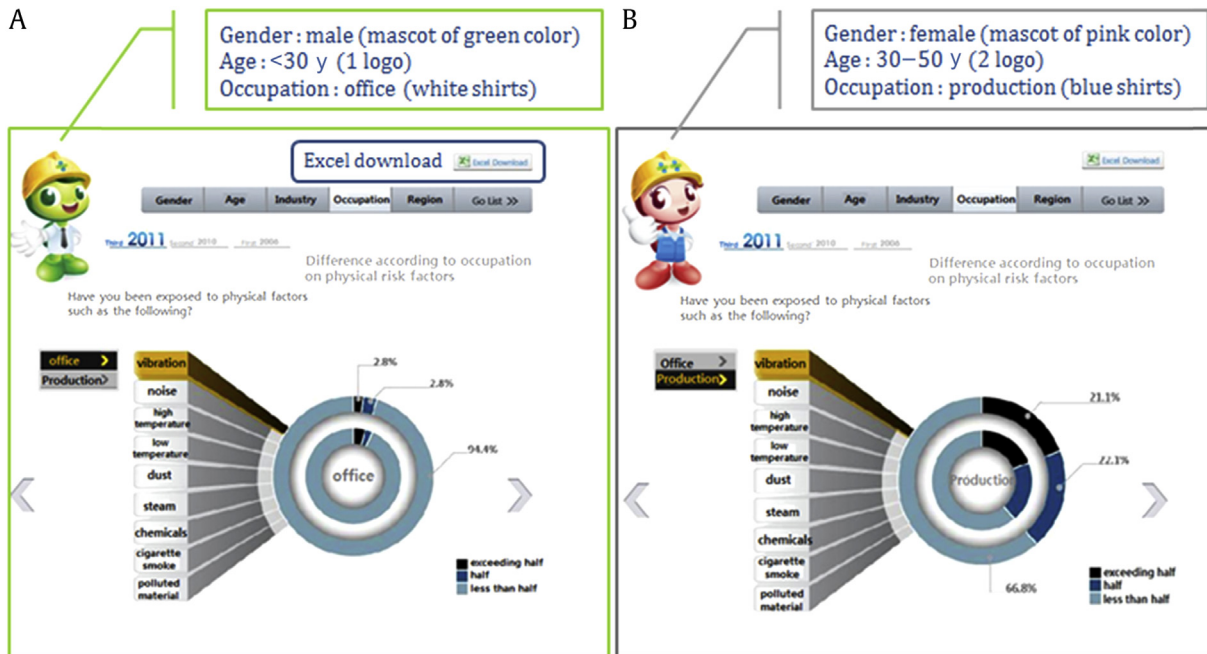

Fig. 6. Dress code of avatars. (A) Male avatar is under 30 years and has an office job, and there is a function to download the analysis data in an excel file. (B) Female avatar is over 30 years and under 50 years, and has a production job.

(or state). Despite providing useful data for users, it has been indicated that this method may not make it easy for general users to interpret the analysis data. To solve problems of data accessibility in the existing database systems, this study applied intuitive visualization or infographics. This is expected to make data more accessible and readable for users. This study has established an analysis system that can reflect users' needs through the database for KWCS, with the objective of managing accumulated basic data systematically. However, our study needs to overcome the following limitations through improvement.

First, it is impossible to understand a yearly trend. In all analyses, the year of KWCS is first selected; then, the analysis results are presented. In other words, it fails to allow one to know if there is any yearly difference or variation through the application of time series analysis. We plan to solve this problem and design an additional system as the future work.

Second, this database has a limitation. For this system, we have drawn a schema and built a database on the basis of limited data: five items for objective sociodemographic characteristics and six items for the results from the KWCS. It fails to respond flexibly to the analysis results of the other data or users' needs. We plan to select additional items with high usage of data and expand the data structure. We also plan to include the European Working Conditions Surveys data to develop them into an analysis system that enables international comparison as well.

Conflicts of interest

The author declares no conflict of interest.

Acknowledgments

References

[1] Lu ML, Nakata A, Park JB, Swanson NG. Workplace psychosocial factors associated with work-related injury absence: a study from a nationally representative sample of Korean workers. Int J Behav Med 2014;21:42−5.

[2] Kim YS, Rhee KY, Oh MJ, Park J. The validity and reliability of the second Korean Working Conditions Survey. Saf Health Work 2013;4:111−6.

[3] Korean Statistical Information Service. Korean Working Conditions Survey for workers (2006∼2011) [Internet]. [cited 2015 Dec]. Available from: http://kosis.kr/statisticsList/statisticsList_01List.jsp?vwcd=MT_OTITLE&parmTabId=M_01_02.

[4] Wikipedia. One-way communication (telecommunications) [Internet]. [cited 2015 Dec]. Available from: http://en.wikipedia.org/wiki/Duplex_(telecommunications).

[5] Wikipedia. Two-way communication [Internet]. [cited 2015 Dec]. Available from: http://en.wikipedia.org/wiki/Two-way_communication.

[6] Ministry of Employment and Labor. Survey report on labor conditions by employment type. Seoul (Korea): Ministry of Employment and Labor; 2011.

[7] Krivitsky PN, Handcock MS, Raftery AE, Hoff PD. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. Soc Networks 2009;31:204−13.

[8] Jansen BJ, Sobel K, Cook G. Classifying ecommerce information sharing behaviour by youths on social networking sites. J Inf Sci 2011;37:120−36.

[9] Kim TH, Lee JH, Lee MK, Jung HM, Kim DW. Suggestion and evaluation on information services in viewpoint of visualization attributes. J Korea Contents Assoc 2011;11:489−99. [in Korean].

[10] Borkin MA, Vo AA, Bylinskii Z, Isola P, Sunkavalli S, Oliva A, Pfister H. What makes a visualization memorable? IEEE Trans Vis Comput Graphics 2013;19:2306−15.

[11] Ilves M, Gizatdinova Y, Surakka V, Vankka E. Head movement and facial expressions as game input. Entertain Comput 2014;5:147−56.

[12] Borghese NA, Pirovano M, Mainetti R, Lanzi PL. IGER: an intelligent game engine for rehabilitation. Berlin (Germany): Springer; 2013. p. 947−50.

[13] Korean Occupational Safety & Health Agency [Internet]. [cited 2015 Dec]. Available from: http://www.kosha.or.kr/.

[14] Lin KW, Ku WC. The application of interpersonal relationship poll to the enhancement of management efficacy. J Soc Sci 2012;8:407−11.

[15] Verklan MT. Johari window: a model for communicating to each other. J Perinat Neonatal Nurs 2007;21:173−4.

[16] Wikipedia. Social learning theory [Internet]. [cited 2015 Dec]. Available from: http://en.wikipedia.org/wiki/Social_learning_theory.

[17] Garmendia E, Stagl S. Public participation for sustainability and social learning: concepts and lessons from three case studies in Europe. Ecol Econ 2010;69:1712−22.

[18] European Working Conditions Surveys [Internet]. [cited 2015 Dec]. Available from: http://www.eurofound.europa.eu/ewco/surveys/index.htm.

[19] Park J, Lee N. First Korean working conditions survey: a comparison between South Korea and EU countries. Ind Health 2009;47:50−4.

[20] Carte TA, Jasperson J, Cornelius ME. Integrating ERD and UML concepts when teaching data modeling. J Inf Syst Educ 2006;17:55−63.